

PCT

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE  
Bureau international



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>3</sup> : C12N 15/00; C07H 21/04; C12P 21/02; C07C 103/52; A01K 39/29; G01N 33/54		A1	(11) Numéro de publication internationale: WO 81/00577 (43) Date de publication internationale: 5 mars 1981 (03.03.81)
(21) Numéro de la demande internationale: PCT/FR80/00133 (22) Date de dépôt international: 29 août 1980 (29.08.80) (31) Numéros des demandes prioritaires: 79/21811 80/09039 (32) Date de priorité: 30 août 1979 (30.08.79) 22 avril 1980 (22.04.80) (33) Pays de priorité: FR		(72) Inventeurs, et (75) Inventeurs/Déposants (US: seulement): GALIBERT, Francis [FR/FR]; 7, avenue Simon Hayen, F-95210 Saint-Gatien (FR). TIOLLAIS, Pierre [FR/FR]; 16, rue de la Glacière, F-75013 Paris (FR). CHARNAY, Patrick [FR/FR]; 1, rue du Parc, F-92100 Boulogne-Billancourt (FR). (74) Mandataires: GUTMANN, Ernest, etc.; Cabinet Plasse-raud, 84, rue d'Amsterdam, F-75009 Paris (FR). (81) Etats désignés: AT, CH, DE, DK, GB, JP, LU, NL, SE, US.	
(71) Déposant (pour tous les Etats désignés sauf US): INSTITUT PASTEUR [FR/FR]; 25-28, rue du Docteur Roux, F-75724 Paris Cedex 15 (FR). INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE [FR/FR]; 101, rue de Tolbiac, F-75645 Paris Cedex 13 (FR).		Publiée Avec rapport de recherche internationale Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues	
(54) Titre: NUCLEOTIDIC SEQUENCE CODING THE SURFACE ANTIGEN OF THE HEPATITIS B VIRUS, VECTOR CONTAINING SAID NUCLEOTIDIC SEQUENCE, PROCESS ALLOWING THE OBTENTION THEREOF AND ANTIGEN OBTAINED THEREBY.			
(54) Titre: SÉQUENCE NUCLEOTIDIQUE CODANT L'ANTIGÈNE DE SURFACE DU VIRUS DE L'HEPATITE B, VECTEUR CONTENANT LADITE SÉQUENCE NUCLEOTIDIQUE, PROCÉDE PERMETTANT SON OBTENTION ET ANTIGÈNE OBTENU			
(57) Abstract Nucleic acid of reduced size and vector containing said nucleotidic sequence of which ABN codes an immunogenic peptidic sequence capable of inducing the generation of antibodies to the virus of viral hepatitis B. It comprises totally or partly the sequence of nucleotides represented in figure 3A. Application to the production by cloning in a bacterium of an immunogenic protein immunising against hepatitis B, or application to the obtention of probes for the diagnosis of the presence of Dane particles in a serum.			
(57) Abrégé Acide nucléique de taille réduite et vecteur contenant ladite séquence nucléotidique dont l'ADN code une séquence peptidique immunogène capable d'induire la production d'anticorps à l'égard du virus de l'hépatite virale B. Il comprend en totalité ou en partie la séquence de nucléotides représentée à la figure 3A. Application à la production par clonage dans une bactérie d'une protéine immunogène vaccinant à l'égard de l'hépatite B ou à l'obtention de sondes pour le diagnostic de la présence de particules de Dane dans un sérum.			
<p>3' -TAC CTC TTG TAG TGT AGT CCG AAG GAT CCG 5' ATG GAG TAC ATC ACA TCA GGA TTT CTA GGA MET GLU ASN ILE THR SER GLY PHE LEU SER 60 GGG GAA GAG CAC AAT GTC CCG CCC AAA TAT GCC CTT CTC GTG TTA CAG GCG GGG TTT TTC PRO LEU LEU VAL LEU GLN ALA GLY PHE PHE 90 AAC AAC TGT TCT TAG GAG TGT TAT GGC GTC TTG TTG ACA GGA ATC CTA ATA CCG CAG LEU LEU THR ARG ILE LEU THR ILE PRO GLY 120 TCA GAT GTG AGC ATC ACC TGA AGA GAG TTA AGT CTA GAC TCG TGG TGG ACT TCT CTC AAT SER LEU ASP SER TRP TRP THR SER LEU ASN 150 AAA GAT CCC CCG TGA TGG GAC ACA GAA CCG TTT CTA GCG GGA ACT ACC GTG TGT ETT GGC PHE LEU GLY GLY THR THR VAL CYS LEU GLY 180 GTT TTA AGC GTC AGG GGT TGG AGG TTA GTG CAA AAT TCG CAG TCC CCA ACC TCC AAT CAC GLN ASN SER GLN SER PRO THR SER ASN HIS 210 AGT GGT TGG AGA ACA GGA GGT TGA ACA GGA TCA CCA ACC TCT TGT GGT CCA ACT TGT CTT SER PRO THR SER CYS PRO PRO THR CYS PRO 240 CCA ATA GCG ACC TAC ACA GAC GCC GCA AAA GGT TAT CCG TGG ATG TGT CTG CCG GGT TTT GLY THR ARG TRP MET CYS LEU ARG ARG PHE</p>			

BEST AVAILABLE COPY

# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AT	Autriche	KP	République populaire démocratique de Corée
AU	Australie	LI	Liechtenstein
BR	Brésil	LU	Luxembourg
CF	République Centrafricaine	MC	Monaco
CG	Congo	MG	Madagascar
CH	Suisse	MW	Malawi
CM	Cameroun	NL	Pays-Bas
DE	Allemagne, République fédérale d'	NO	Norvège
DK	Danemark	RO	Roumanie
FI	Finlande	SE	Suède
FR	France	SN	Sénégal
GA	Gabon	SU	Union soviétique
GB	Royaume-Uni	TD	Tchad
HU	Hongrie	TG	Togo
JP	Japon	US	Etats-Unis d'Amérique

" SEQUENCE NUCLEOTIDIQUE CODANT L'ANTIGENE DE SURFACE DU VIRUS DE L'HEPATITE B, VECTEUR CONTENANT LADITE SEQUENCE NUCLEOTIDIQUE, PROCEDE PERMETTANT SON OBTENTION ET ANTIGENE OBTENU "

L'invention est relative à un acide nucléique comprenant une séquence nucléotidique capable de coder une séquence peptidique immunogène correspondant à l'antigène de surface du virus de l'hépatite virale B, et aux poly-  
5 peptides et peptides obtenus.

Elle concerne également un procédé permettant d'obtenir un tel acide nucléique.

L'hépatite B est une maladie virale fréquente tout particulièrement en Afrique Tropicale, en Asie du  
10 Sud-Est et en Extrême-Orient.

L'agent étiologique est un virus (HBV) ou particule de Dane, comprenant une enveloppe (antigène Australia ou antigène HBs), une capside (antigène HBc), une polymérase endogène et une molécule d'ADN circulaire  
15 partiellement simple brin ; le brin le plus long de cette molécule d'ADN comporte près de 3 200 nucléotides (SUMMERS J., O'CONNELL A. et MILLMAN I. (1975) Proc. Nat. Acad. Sci. USA 72, 4 597-4 601).

L'ADN polymérase endogène peut être utilisée  
20 pour réparer in vitro le brin le plus court (T. A. LANDERS, H. B. GREENBERT et J. S. ROBINSON, J. VIROL., 23, 1977, p. 368-376).

L'analyse électrophorétique des protéines de l'enveloppe a montré la présence de 2 à 7 polypeptides  
25 dont les principaux sont appelés : polypeptide I et polypeptide II (PETERSON D. L., ROBERTS I. M. et VYAS G. N. (1977) Proc. Nat. Acad. Sci., USA, 74, 1 530-1 534, et PETERSON D. L., CHIEN D. Y., VYAS G. N., NITECKI D. et BOND H. (1978) In Viral Hepatitis, ed. G. VYAS, S. COHEN  
30 et R. SCHMID, The Franklin Institute Press, Philadelphie, 569-573).



Le polypeptide I a un poids moléculaire de 22 000 à 26 000 daltons. Le polypeptide II est glycosilé et a un poids moléculaire de 28 000 à 30 000 daltons. La composition en acides aminés de ces deux polypeptides est très semblable; les séquences que forment, respectivement, leurs 15 premiers amino-acides (à partir de l'extrémité N-terminale) et leurs 3 derniers amino-acides sont identiques, de sorte que l'hypothèse a été formulée que le polypeptide II pourrait ne différer du polypeptide I que par une glycosilation.

L'étude du virus est extrêmement difficile dans la mesure où l'on ne dispose d'aucun système de culture cellulaire permettant la propagation du virus. Cette difficulté a déjà en partie été contournée, plus particulièrement en ce qui concerne le sérotype ayw. L'ADN entier (génome) du virus a été identifié et cloné, notamment dans E. coli, après son insertion préalable dans le site unique EcoRI d'un vecteur  $\lambda$ gt.WES.  $\lambda$ B, selon la technique par FRITSCH A., POURCEL C., CHARNAY P. et TIOLLAIS P. (1978) C. R. Acad. de Paris, 287, 1 453-1 456).

A ce jour, la séquence des polypeptides I et II eux-mêmes, la localisation dans l'ADN viral de la séquence codant ces peptides n'ont pas été réalisées.

L'invention a pour but l'obtention d'une séquence d'ADN beaucoup plus petite que l'ADN viral lui-même, contenant la séquence apte à coder la séquence peptidique douée de propriétés immunogènes permettant, lorsqu'elle est introduite dans l'organisme d'un hôte vivant, d'induire la fabrication par ce dernier d'anticorps susceptibles de protéger ultérieurement ce même hôte à l'égard du virus de l'hépatite virale B, notamment lorsque celui-ci se trouve à l'état virulent.

L'invention découle non seulement de l'analyse nucléotidique complète du génome de la particule de Dane



à laquelle ont procédé les inventeurs, mais à l'idée qu'ils ont eue pour identifier le gène codant (ci-après dénommé "gène S") des susdits polypeptides, de rechercher dans la structure nucléotidique complète ainsi préétablie du gé-

5 nome de la particule de Dane, celles des séquences de nucléotides susceptibles de coder les séquences peptidiques proximales et terminales connues de ces polypeptides.

On rappellera que PETERSON et collaborateurs ont rapporté, notamment dans les articles dont les références ont été rappelées plus haut, que la séquence proximale (premier amino-acide N-terminal) des 15 premiers amino-acides s'analysait en principe comme suit :

10 met glu asn ile thr ser gly phe leu gly pro leu leu val ser et que la séquence terminale de ces mêmes polypeptides

15 (dernier amino-acide C-terminal) était la suivante :

val tyr ile

La figure 1 est une carte schématique du génome de la particule de Dane. Celui-ci comporte deux brins  $b_1$  et  $b_2$  ; le plus court d'entre eux ( $b_2$ ) étant normalement

20 dépourvu de la partie représentée par une ligne en trait interrompu dans le dessin.

Il est connu que cet ADN ne comporte qu'un seul site EcoRI.

La flèche  $f_1$  donne le sens de la numérotation des nucléotides dont est composé le brin le plus long  $b_1$ , et la flèche  $f_2$  donne le sens de la transcription de l'ADN du virus, notamment par la machinerie cellulaire des cellules envahies par le virus de l'hépatite B, pour ce qui

25 concerne l'expression du gène S.

30 Le site EcoRI peut donc être numéroté 0 ou, comme on l'a maintenant déterminé de façon plus exacte pour celui des virus de l'hépatite B appartenant au sérotype ayw, 3 182.

Le cercle en trait plein intérieur e donne l'é-

35 chelle en % de longueur de l'ADN et permet de préciser les



positions de certaines de ses parties.

Les nombres 3', 5' et 5', 3' à la partie inférieure de la carte visent les extrémités terminales porteuses de mêmes nombres dans les représentations conventionnelles des extrémités des chaînes d'acide nucléique.

Selon l'invention, il a été montré que le "gène S" constituait essentiellement le fragment du brin le plus long  $b_1$  situé entre les positions 73,6 et 95,1 de la carte schématique de la figure 1. Les abréviations "Start" et "Stop" représentent les points d'initiation et d'arrêt de la transcription du "gène S".

Les figures 2A, 2B, 2C sont représentatives de la partie terminale du susdit génome, comprise notamment entre les positions 60,4 et 100 (en % de longueur de l'ADN). Chacune des lettres figurant dans la figure 2 correspond de façon conventionnelle à l'un des 4 nucléotides de base de l'ADN :

A : Adénine

G : Guanine

T : Thymine

C : Cytosine

Les lignes inférieures, dans chaque paire de lignes dont sont constituées les figures 2A, 2B, 2C, correspondent à l'acide nucléique correspondant à la chaîne nucléotidique  $b_2$ .

La technique d'analyse utilisée pour établir la carte plus détaillée dont témoignent les figures 2A, 2B, 2C, sera brièvement rappelée ci-après.

La caractérisation de la séquence nucléotidique du "gène S" telle qu'elle est proposée dans le cadre de la présente invention, et dont les extrémités proximale  $p$  "S" et terminale  $t$  "S" sont indiquées dans les figures 2A, 2B, 2C, résulte de la constatation que :

- les 14 premiers triplets (dans le sens de la lecture  $f_2$ ) à partir du nucléotide numéroté 3 030 vis-à-vis de l'extrémité terminale EcoRI, sont respectivement

FEUILLE DE REMPLACEMENT



capables de coder les 14 premiers amino-acides de la séquence proximale des 15 premiers amino-acides des susdits polypeptides,

- les 4 derniers triplets GTA TAC ATT TAA lus sur la chaîne complémentaire  $b_2$  à la chaîne transcrite  $b_1$  correspondent respectivement aux 3 amino-acides terminaux des susdits polypeptides et à un codon d'arrêt ;

- cette séquence de nucléotides (678 nucléotides) ne comprend aucun codon d'arrêt, du moins lorsque l'on adopte le cadre de lecture impliquant que le premier triplet "lu" sur l'ADN par la machinerie cellulaire soit AUG, (correspondant sur le brin complémentaire à ATG) ;

- la traduction complète de l'information génétique commençant avec le codon initial ATG conduit à un polypeptide théorique de 226 amino-acides, ayant un poids moléculaire de 25 422 daltons.

La structure nucléotidique du "gène S" ainsi que la chaîne polypeptidique résultant de la traduction du "gène S" sont représentées aux figures 3A, 3B, 3C.

Ces valeurs sont tout à fait conformes aux données analytiques qui résultent de la mobilité électrophorétique du polypeptide I sur des gels de polyacrylamide qui ont déjà été décrits par les auteurs précédents (références 9-12 selon la bibliographie figurant à la fin de la description de la présente demande de brevet).

La différence observée au niveau du 15ème amino-acide de la séquence peptidique proximale du polypeptide I : leucine selon les cartes des figures 2A, 2B, 2C et 3A, 3B, 3C mentionnées ci-dessus, et non sérine selon les constatations des susdits auteurs, peut peut-être être attribuée au fait que ces auteurs ont travaillé avec un variant génétique différent de celui ayant fait l'objet de la présente étude. On notera que la différence peut d'ailleurs être rapportée à la substitution d'un seul nucléotide dans le triplet concerné "TTA" dans le "gène S" particulier représenté dans les cartes des figures 2A, 2B, 2C et 3A, 3B, 3C, au lieu de "TCA", l'un des

FEUILLE DE REMPLACEMENT



triplets susceptibles d'être traduits en sérine.

L'invention concerne donc plus particulièrement les fragments d'acide nucléique qui peuvent être excisés de l'ADN de la particule de Dane, ces fragments étant plus particulièrement caractérisés en ce qu'ils contiennent la partie du "gène S" susceptible de coder la partie de la protéine de l'enveloppe du virus qui est responsable des propriétés immunologiques du virus de l'hépatite B.

A ce titre, l'invention concerne donc un acide nucléique comportant au plus de l'ordre de 1 000-1 100 nucléotides, plus particulièrement caractérisé en ce qu'il est apte à coder une séquence peptidique immunogène, elle-même apte à induire in vivo la production d'anticorps actifs à l'égard du virus de l'hépatite B, cette séquence peptidique contenant essentiellement la structure représentée aux figures 3A, 3B, 3C, ou toute séquence peptidique ayant des propriétés immunogènes équivalentes.

L'invention concerne également un vecteur en vue de l'expression de ladite séquence nucléotidique dans un micro-organisme ou dans des cellules eucaryotes à condition que la fusion génétique ait été effectuée en conservant la phase de lecture du "gène S".

Les séquences nucléotidiques utilisées conformes à l'invention présentent l'une par rapport à l'autre une variabilité conduisant, lors de leur expression, à la formation de déterminants variant selon le sous-type du virus de l'hépatite B (sous-types d, w, y, r du groupe a).

Pour l'une des séquences peptidiques représentées aux figures 3A, 3B, 3C, on observera que le premier acide aminé de la séquence susdite : méthionine, est N-terminal et que l'acide-amino de l'extrémité opposée : isoleucine, est C-terminal.

L'invention concerne encore, plus particulièrement, la séquence nucléotidique représentée aux figures 4A, 4B, codant la séquence peptidique telle qu'elle résulte de la

FEUILLE DE REMPLACEMENT





figure 5 ou toute séquence peptidique analogue douée de propriétés immunogènes équivalentes.

Il va de soi que par "séquence peptidique équivalente", mentionnée ci-dessus, il faut entendre toute séquence peptidique dans laquelle certaines parties peuvent n'être pas rigoureusement identiques aux parties correspondantes de la séquence peptidique représentée dans les figures 3A, 3B, 3C et 5, ces variations pouvant être dues à des mutations locales n'affectant pas le caractère immunogène général de la protéine ou à des modifications de structure tenant aux différents sérotypes sous lesquels les protéines du genre en question sont susceptibles de se présenter (notamment sérotypes adw, adr et ayr).

L'invention concerne plus particulièrement la séquence nucléotidique contenant la séquence peptidique telle qu'elle est représentée à la figure 6 ou toute séquence peptidique analogue douée de propriétés immunogènes équivalentes.

L'invention concerne plus particulièrement encore les séquences peptidiques suivantes :

Alanine-Glutamine-Glycine-Thréonine-Sérine  
Thréonine-Alanine-Glutamine-Glycine-Thréonine-Sérine  
Thréonine-Thréonine-Alanine-Glutamine-Glycine-Thréonine-Sérine

Dans le premier peptide sus-indiqué l'extrémité alanine est N-terminale et l'extrémité sérine est C-terminale.

Dans le deuxième et troisième peptides sus-indiqués, l'extrémité thréonine est N-terminale et l'extrémité sérine est C-terminale.

A titre d'exemple, on peut notamment préparer le pentapeptide en partant de la sérine C-terminale à laquelle on accroche la thréonine par la méthode de Castro décrite dans Tétrahédon Letters, 1975, n°14, page 1219-1222. Ensuite les acides aminés glycine, glutamine, alanine sont rajoutés par la méthode dite de l'anhydride mixte répétée (méthode rema) décrite par Beierman dans Chemistry and Biology of Peptides, Ed. J. Meienhofer, Ann. Arbor

FEUILLE DE REMPLACEMENT



Science Publ., Ann. Arb. Mich. 351 (1972).

L'invention concerne également les produits résultant de la fixation du pentapeptide sur une molécule porteuse plus grosse, notamment de type polypeptide ou protéine, 5 les compositions contenant ce pentapeptide en produits de fixation, notamment en association avec un véhicule pharmaceutiquement acceptable, et plus particulièrement des vaccins contre l'hépatite B. Ces véhicules pharmaceutiques sont appropriés, de façon en soi classique, au mode d'administration choisi, notamment par voie orale, parentérale, rectale ou 10 par nébulisation sur des muqueuses, notamment nasales.

L'hexapeptide et le polypeptide à 7 acides aminés peuvent être synthétisés par des techniques de synthèse peptidique classiques.

Ces peptides sont, selon la présente invention, 15 tenus pour être le site antigénique des polypeptides de plus grande taille dont question plus haut et responsables du pouvoir vaccinant de l'enveloppe virale (Journal of Biol. Stand. 1976, 4, 295-304, RAO et VYAS "Biochemical Characterization of Hepatitis B Surface Antigen in Relation 20 to Serologic Activity").

L'invention concerne également encore les fragments d'ADN susceptibles de coder la production de tels pentapeptide, hexapeptide et polypeptide à 7 acides aminés. Il s'agit :

25 - pour le pentapeptide, notamment du polynucléotide de formule:

5' GCT CAA GGA ACC TCT 3'

3' GGA GTT CCT TGG AGA 5'

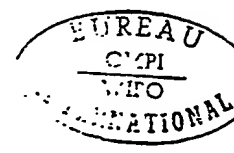
- pour l'hexapeptide, notamment du polynucléotide de de formule : 30

5' ACT GCT CAA GGA ACC TCT 3'

3' TGA CGA GTT CCT TGG AGA 5'

- pour le polypeptide à 7 acides aminés du polynucléotide de formule :

35 5' ACT ACT GCT CAA GGA ACC TCT 3'



3' TGA TGA CGA GTT CCT TGG AGA 5'

ou dans chacun des trois cas, du polynucléotide complémentaire relatif aux trois polynucléotides respectifs précédents ou encore tout polynucléotide dans lequel chacun  
5 des triplets peut être remplacé par tout triplet analogue capable de coder la production du même amino-acide.

L'acide nucléique selon l'invention peut encore être caractérisé en ce qu'il comporte au moins l'un des deux brins mutuellement complémentaire d'une séquence  
10 d'ADN telle que représentée dans les figures 4A, 4B (dans lesquelles figurent également les nombres correspondant aux positions des premiers nucléotides de chacun des fragments successifs de 10 nucléotides représentés vis-à-vis de la position EcoRI non représentée dans la figure : il va de  
15 soi que ces nombres n'interviennent pas au niveau de la caractérisation de la séquence nucléotidique du genre en question). Ce fragment d'ADN est délimité par deux sites HincII.

On appréciera que cette séquence nucléotidique  
20 correspond à l'information génétique dont la traduction conduit à la séquence peptidique représentée à la figure 5.

L'invention concerne naturellement les séquences nucléotidiques équivalentes à simple brin ou à double brin, dont notamment le brin ayant la structure qui découle de  
25 la succession des lignes inférieures des figures 4A, 4B, l'ADN à double brin correspondant, ou les ARN messagers correspondants, notamment celui représenté par les enchaînements complémentaires de nucléotides constitués par les lignes inférieures des paires de lignes des figures 4A, 4B (sens  
30 de la flèche  $f_2$ ).

De même entrent dans le champ de l'invention les chaînes nucléotidiques qui se différencient des précédentes par certains triplets ou petites séquences de triplets, dans la mesure où ces séquences nucléotidiques restent  
35 aptes à coder un polypeptide conservant les activités

FEUILLE DE REMPLACEMENT



immunogènes caractéristiques des virus de l'hépatite virale B. D'une façon générale, il s'agit de chaînes de nucléotides qui, le cas échéant, après dénaturation de l'ADN à double brin pour produire les acides nucléiques à simple  
5 brin correspondants, restent susceptibles de s'hybrider sur au moins environ 90 % de leur longueur avec l'un des brins de l'ADN des figures 4A, 4B.

Des acides nucléiques préférés selon l'invention sont encore ceux qui peuvent être excisés à partir de  
10 l'ADN de l'hépatite virale et qui, lorsqu'ils sont à double brin, sont caractérisés par l'existence à l'un de leurs bouts d'une extrémité HincII, HhaI, AvaI ou EcoRI et à leur autre bout par une extrémité AvaIII, HincII ou HhaI.

15 Les positions de ces diverses extrémités vis-à-vis du site EcoRI sont schématisées dans les figures 2A, 2B, 2C.

L'acide nucléique selon l'invention est destiné à l'incorporation dans un vecteur permettant son expression dans une bactérie et dans les cellules eucaryotes,  
20 notamment en vue de la production d'une protéine ou d'un peptide susceptible d'induire dans l'organisme d'un hôte vivant la production d'anticorps actifs contre le virus de l'hépatite virale B. La protéine ou le peptide résultant de la traduction de la séquence nucléotidique selon  
25 l'invention peuvent être utilisés comme agent vaccinant ou comme agent servant pour le diagnostic.

L'acide nucléique selon l'invention peut également être utilisé comme sonde pour dépister la présence ou non dans des échantillons de sang ou de sérum d'épreuve,  
30 de particule de Dane, d'antigène HBs ou de fragments de celui-ci, etc. (par technique classique d'hybridation ADN-ADN).

D'autres caractéristiques de l'invention résulteront encore de la description brève qui suit des techniques d'analyse, d'identification et d'obtention de  
35

FEUILLE DE REMPLACEMENT



fragments d'ADN conformes à l'invention. Référence sera naturellement faite aux dessins dont les figures ont déjà été prises en considération dans ce qui précède. Les chiffres ou nombres entre parenthèses correspondent aux références de la bibliographie annexée à la présente description.

L'invention concerne aussi des vecteurs particuliers permettant l'expression des séquences nucléotidiques décrites ci-dessus, notamment sous la forme d'une protéine hybride dans laquelle un fragment protéique présentant les caractères immunologiques de HBsAg jouxte une molécule porteuse conférant à l'ensemble des propriétés immunogènes ou immunoréactives, susceptibles d'induire la production d'anticorps protecteurs à l'égard de l'infection virale dans l'organisme de l'hôte dans lequel cette protéine a au préalable été introduite.

En particulier, l'invention concerne un vecteur - phage ou plasmide - contenant au moins une partie de l'opéron lactose, plus particulièrement le promoteur et le gène Z de cet opéron, ce vecteur étant caractérisé en ce qu'il est modifié pour l'insertion, en phase, dans un site approprié du gène Z, tel que le site EcoRI de l'un quelconque des fragments d'ADN du brevet principal, notamment ceux contenant la plus grande partie du "gène S". Elle concerne également ceux de ces vecteurs modifiés, dans lesquels une partie au moins du fragment d'ADN codant pour la plus grande partie de la  $\beta$ -galactosidase serait remplacée par un fragment d'ADN apte à coder pour toute autre molécule porteuse non immunogène, ou dont les éventuelles propriétés immunologiques, si celles-ci existent, n'interfèrent pas avec celles de la partie peptidique présentant les propriétés immunologiques de HBsAg, par exemple essentiellement celle qui s'étend dans la direction de la lecture à partir de son site HhaI.

L'invention est également plus particulièrement



relative à une protéine hybride caractérisée en ce qu'elle contient une séquence polypeptidique présentant les propriétés immunologiques spécifiques de HBsAg, jouxtant une séquence polypeptidique constituée par la majeure partie de la  $\beta$ -galactosidase, qui joue le rôle de protéine-porteuse.

L'invention ne s'étend pas seulement à cette molécule hybride particulière, dont le rôle essentiel est de constituer un modèle d'une protéine construite selon les techniques du génie génétique et douée des propriétés immunogènes et immunoréactives caractéristiques de l'antigène HBsAg, mais également à toute autre protéine hybride dans laquelle tout ou partie de la partie  $\beta$ -galactosidase peut être remplacée par toute autre molécule porteuse non immunogène, ou dont les éventuelles propriétés immunologiques, si celles-ci existent, n'interfèrent pas avec celles de la partie peptidique présentant les propriétés immunologiques de HBsAg.

D'autres caractéristiques de l'invention apparaîtront encore au cours de la description d'exemples préférés, en combinaison avec les dessins dans lesquels :

- les figures 1a et 1h illustrent schématiquement les étapes successives de la fabrication d'un vecteur du type plasmide incorporant un fragment de HBV DNA,

- les figures 2a à 2c illustrent schématiquement les structures initiales du vecteur utilisé (figure 2a) final du vecteur modifié obtenu (figure 2b) et celle de la protéine hybride résultant de l'expression de ce vecteur modifié dans E. coli (figure 2c).

#### A - SEQUENCES NUCLEOTIDIQUES

##### 30 Les produits et méthodes utilisés

##### - Les enzymes et les produits chimiques utilisés

Les enzymes de restriction utilisées : BamHI, HhaI, HincII, HaeIII, XbaI, MboI, HinfI, HpaII, XhoI, sont celles fabriquées par BIOLABS. On a utilisé l'ADN-polymérase I de BOEHRINGER. La phosphatase alcaline bactérienne



et la polynucléotide-kinase ont été fournies par P. L. BIOCHEMICALS. Les agents chimiques étaient les suivants :

- 5                   . Diméthyl sulfate (ALDRICH),
- . Hydrazine (EASTMAN KODAK),
- . Acrylamide et bis-acrylamide (2 fois cristallisés - SERVA),
- . Dideoxy nucléotide triphosphates et deoxynucléotide triphosphates (P. L.
- 10 BIOCHEMICALS),

                  . Pipéridine (MERCK) redistillée sous vide.

- Préparation d'ADN HBV

Le génome HBV entier (sous-type ayw) a été cloné dans E. coli en mettant en jeu le site unique de restriction EcoRI du vecteur  $\lambda$ gt.WES.  $\lambda$ B (14). L'ADN cloné est dénommé ci-après "Eco HBV DNA".

Le bactériophage recombinant a été amplifié en boîte de Pétri sur Agar et on a extrait l'ADN recherché de façon en soi connue. Après digestion de l'ADN par l'enzyme de restriction EcoRI, on a purifié la séquence Eco HBV DNA par ultracentrifugation, dans un gradient de sucrose, selon la technique décrite dans les références bibliographiques (16, 17).

- Préparation de fragments d'ADN marqués 5',<sup>32</sup>P

25                   10 à 20 picomoles d'Eco HBV DNA ont été complètement hydrolysés par les différentes enzymes de restriction, dans les conditions recommandées par le fabricant. Les fragments d'ADN ont été déphosphorylés par la phosphatase alcaline, celle-ci ayant ensuite été inactivée

30 par un traitement alcalin. L'ADN a alors été précipité à l'éthanol, selon la technique décrite dans l'article (18). Après redissolution dans un tampon à base de spermidine, les ADN ont été marqués à leurs extrémités 5' avec un ATP [ $\lambda^{32}$ P] (3 000 Ci/mM fabriqué par NEW ENGLAND

35 NUCLEAR) et avec de la polynucléotide-kinase (selon la



technique visée à l'article (19).

Les fragments d'ADN de restriction ont été séparés par électrophorèse sur gel de polyacrylamide, puis élués. Les extrémités marquées ont fait l'objet de ségrégations par électrophorèse sur gel de polyacrylamide de façon connue en soi, après restriction avec une autre enzyme ou par dénaturation des fragments d'ADN du genre en question.

10     - Détermination de la structure des séquences de nucléotide de l'ADN

La structure primaire des fragments d'ADN à double brin ou simple brin a été déterminée essentiellement selon la technique décrite par MAXAM et GILBERT (19). L'on a aussi eu recours à la méthode des inhibiteurs terminaux de chaînes décrites par SANGER et al. (20) et adaptée par MAAT et SMITH (21), pour ce qui est des fragments à double brin marqués à l'une de leurs extrémités 5'.

Les produits de réaction chimique et enzymatique ont été analysés par électrophorèse dans des gels de séquence d'acrylamide à 8,16 ou 25 %, de 1 mm d'épaisseur.

20     Les techniques et les résultats de l'analyse

Afin de déterminer si le génome HBV est capable de coder les polypeptides I et II, tous les fragments HaeIII (sites de restriction HaeIII du génome HBV représentés sur la figure 1 par de petites flèches) ont été marqués à leurs extrémités 5'. Des parties substantielles de leurs structures primaires ont été déterminées par la méthode de MAXAM et GILBERT. Les séquences de nucléotides susceptibles de coder les séquences d'acide aminé proximales et terminales des polypeptides I et II ont été localisées dans les fragments HaeIII E et HaeIII F, préalablement localisés sur la carte de restriction du génome HBV (selon la technique décrite dans la référence (17)). Ce sont ces séquences de nucléotides qui ont été considérées comme consistant en les extrémités du "gène S"





occupant elles-mêmes les positions 73,6 et 95,1 vis-à-vis du site de restriction EcoRI (figure 1) pour les raisons déjà indiquées.

La séquence de nucléotides entre ces deux positions a été analysée en ayant recours aux techniques chimiques connues, notamment par la méthode de dégradation chimique à l'hydrazine diméthyl sulfate et la méthode des terminaisons de chaînes. On a eu recours, parmi les réactions chimiques variées proposées par MAXAM et GILBERT à une dépurination partielle par l'acide formique et au clivage par la pipéridine, méthodes qui donnent des bandes d'intensité égale sur les autoradiogrammes pour les fragments terminés par une guanine et une adénine. On a également utilisé les réactions à l'hydrazine suivies d'un clivage à la pipéridine, pour obtenir des bandes d'égale intensité, pour les nucléotides cytosine et thymidine : le fractionnement électrophorétique des produits de ces deux réactions donne pour toutes les bases une tache dans l'une ou l'autre des colonnes de gel utilisées. Cette procédure facilite la lecture de l'autoradiogramme du gel. La réaction à l'hydrazine en présence de chlorure de sodium spécifique pour la cytosine permet de distinguer ce nucléotide de la thymidine et la réaction au diméthyl sulfate suivie d'un clivage par la pipéridine spécifique de la guanine permet de distinguer ce dernier nucléotide de l'adénine.

De façon à s'assurer du plus grand possible degré de précision, des séquences distinctes de nucléotides formant différents fragments se chevauchant mutuellement ont été produites par hydrolyse d'Eco HBV DNA par diverses enzymes de restriction :

BamHI, HinfI, HpaII, HaeIII et HincII.

De cette façon l'analyse de chacun des sites de restriction utilisés comme points de départ de premiers fragments étudiés a été confirmée par l'analyse des



fragments distincts dans lesquels les sites de restriction des premiers fragments se trouvaient compris entre les nouvelles extrémités de ces fragments distincts.

Le "gène S" représenté aux figures 3A, 3B, 3C, qui commence par le codon d'initiation ATG, comprend 227 triplets, dont un codon d'arrêt TAA. Les 3 codons correspondant aux 3 acides aminés de l'extrémité carboxy terminale du polypeptide correspondant sont situés dans le même cadre de lecture, immédiatement avant le codon d'arrêt TAA. L'un des deux autres cadres de lecture (respectivement décalés vis-à-vis du précédent de 1 et 2 nucléotides) est également dépourvu de codon d'arrêt, mais code une protéine tout à fait différente des polypeptides I et II sus-mentionnés. Le troisième cadre de lecture comprend 10 codons d'arrêt (5 TAG, 4 TGA, 1 TAA). Sur l'autre brin d'ADN, les trois cadres de lecture se trouvent respectivement fermés par 11, 11, et 6 codons d'arrêt distribués le long de la séquence d'ADN.

Comme on l'a déjà indiqué plus haut, la traduction complète de l'information génétique débutant par le codon d'initiation ATG conduit à un polypeptide théorique de 226 amino-acides correspondant à un poids moléculaire de 25 422 daltons.

Il est intéressant de souligner que la séquence nucléotidique correspondant au "gène S" devrait normalement être lue entièrement au cours de la traduction.

Sont également parties de l'invention des chaînes de nucléotides du type du "gène S" ci-dessus décrit, qui comporte de petites séquences supplémentaires pouvant contenir jusqu'à une centaine de nucléotides ou qui au contraire peuvent en être dépourvues, sans que soit pour autant altérée l'information génétique correspondante (22, 23).

Les différents fragments de l'invention qui ont été définis plus haut peuvent être obtenus à partir de la séquence d'ADN dite Eco HBV DNA, en ayant recours aux

FEUILLE DE REMPLACEMENT



enzymes de restriction correspondantes et aux techniques de fractionnement connues des fragments d'ADN, notamment sur gel de polyacrylamide et en mettant à profit leurs migrations sur des distances qui sont fonction de leurs poids moléculaires. Ainsi peut-on par exemple obtenir le fragment dont l'une des extrémités est délimitée par un site EcoRI et l'autre par un site AvaIII en opérant une restriction Eco HBV DNA par l'enzyme AvaIII, le fragment recherché consistant dans le plus petit fragment obtenu (un seul site AvaIII dans Eco HBV DNA).

Le fragment délimité par des extrémités opposées EcoRI et HhaI est obtenu par hydrolyse d'Eco HBV DNA par EcoRI d'abord, puis par hydrolyse partielle par l'enzyme de restriction HhaI. L'on récupère alors parmi les produits de restriction celui qui contient le site AvaIII.

Ces techniques de restriction n'ont évidemment été proposées qu'à titre d'exemple, étant bien entendu que le spécialiste est à même de déterminer les ordres de traitement par les enzymes de restriction pour isoler, à partir notamment d'Eco HBV DNA, les fragments ayant les extrémités de restriction utiles.

A toutes fins utiles, on rappellera que ces opérations de restriction peuvent être réalisées au sein d'un tampon Tris 10 mM ; pH 7,8 ;  $MgCl_2$  6 mM ;  $\beta$ -mercapto-éthanol 6 mM, le même milieu contenant en outre et de préférence 50 mM de NaCl lorsque EcoRI est utilisé.

Comme il a déjà été dit, l'invention concerne l'utilisation des fragments d'ADN décrits à titre de sonde permettant le diagnostic de la présence dans un sérum de particules de Dane ou particules dérivées de la précédente, porteuses d'un ADN susceptible de coder une protéine immunogène caractéristique de l'hépatite B.

L'ADN selon l'invention peut également être incorporé à un vecteur permettant, à condition que l'incorporation ait été réalisée en phase, l'expression de cet



ADN dans une bactérie ou autre micro-organisme, ou dans des cellules eucaryotes.

B - VECTEURS CONTENANT UNE SEQUENCE NUCLEOTIDIQUE DE L'ANTIGENE HBs

5     Construction d'un bactériophage recombinant  $\lambda$ lac HBs-1

Les produits au niveau des différentes étapes de cette construction sont visés dans les figures 1a et 1h. Ils sont également visés par les numéros 1a à 1h.

On a indiqué dans la figure 1a les positions du  
10 "gène S" et de certains sites d'enzyme de restriction.

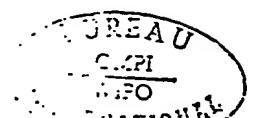
Après traitement de DNA + HBV avec l'enzyme de restriction HhaI, on sépare un fragment d'ADN (1b) contenant 1 084 paires de bases, et plus particulièrement la totalité du "gène S", par électrophorèse sur gel d'agarose  
15 et électro-élution (figure 1b). On prépare, à partir de ce sous-fragment, traité au préalable par l'endonucléase S1, un sous-fragment (1c) (figure 1c), résultant de l'allongement du sous-fragment (1b) à ses extrémités, par des éléments d'ADN dénommés "EcoRI linkers" de formule :

20             5' GGAATTCC  
              CCTTAAGG 3'

Le fragment obtenu est, après formation des extrémités cohésives EcoRI, cloné dans le plasmide pBR322.

Le plasmide obtenu dénommé ci-après pBRHBs (figure 1d), ne contient qu'un seul site de restriction XbaI  
25 situé à proximité de la tête du "gène S".

Par digestion du plasmide recombinant pBRHBs avec un mélange d'enzymes EcoRI et XbaI on produit un fragment d'ADN comprenant approximativement 980 paires  
30 de bases et comportant la majeure partie du "gène S" (figure 1e). Ce fragment est séparé et purifié par électrophorèse sur un gel d'agarose. Le fragment obtenu est à nouveau traité avec l'endonucléase S1, puis à nouveau pourvu d'extrémités EcoRI au moyen des susdits "EcoRI  
35 linkers", puis soumis à un traitement avec l'endonucléase



EcoRI pour reformer les extrémités cohésives correspondantes. Le fragment de la figure 1e qui comprend environ 980 paires de bases est alors inséré par fusion in vitro dans le site EcoRI du plasmide pBR322, pour former le  
5 plasmide pXbaHBs (figure 1f). Ce plasmide est cloné de façon usuelle comme le plasmide pBR322.

Plusieurs clones ont été obtenus.

On extrait et purifie, après traitement avec EcoRI des ADN de trois de ces clones, pXbaHBs-1, pXbaHBs-2  
10 pXbaHBs-3 (figure 1g), les fragments appelés ci-après "fragments HBs" (figure 1h).

Les séquences nucléotidiques des extrémités des susdits fragments (normalement obtenues à l'intérieur du "gène S") sont déterminées en ayant recours à la procédure  
15 décrite par MAXAM et GILBERT (Proc. Nat. Acad. Sci. USA 74, 560-564 (1977)). Ces déterminations ont révélé que les séquences des nucléotides des extrémités terminales, correspondant au "gène S" n'étaient pas identiques dans les trois clones (figure 1g), les différences sont vraisem-  
20 blablement dues à des hétérogénéités produites au cours de la digestion par l'endonucléase S1.

Les deux fragments provenant de pXbaHBs-1 et pXbaHBs-2 sont insérés par fusion in vitro dans le génome du bactériophage  $\lambda$ plac 5-1 (21), lequel n'a qu'un seul  
25 site EcoRI situé à proximité de l'extrémité du gène lac Z. Du fait du cadre de lecture du gène lac Z, tel qu'il peut être déduit de la séquence d'acides aminés de la  $\beta$ -galactosidase (23), on constate -et l'expérience le confirmera- que l'insertion du fragment HBs de pXbaHBs-1  
30 dans le site EcoRI du gène lac Z de  $\lambda$ plac 5-1 doit conduire à la conservation de la phase de lecture adéquate du "gène S". Au contraire, l'insertion du fragment HBs de pXbaHBs-2 devait se révéler ne pas pouvoir être insérée dans le précédent vecteur avec conservation du cadre de  
35 lecture approprié. Il a néanmoins été utilisé comme



témoin dans des expériences postérieures.

Ces opérations ont été réalisées en ayant recours à des techniques connues. En particulier les "fragments HBs" de pXbaHBs-1, pXbaHBs-2 ont été insérés au moyen d'une ligase dans l'ADN de  $\lambda$ plac 5-1 qui avaient au préalable été clivés par EcoRI. Les mélanges des fragments de ADN obtenus ont été ensuite utilisés pour transfecter la souche C600RecBC  $rk^-mk^-$  de E. coli. Les clones de bactériophage devenus  $lac^-$  du fait de l'insertion des fragments HBs dans les sites EcoRI du gène lacZ sont amplifiés et purifiés selon la méthode décrite en (21).

Les ADN des différents bactériophages ont été extraits et les orientations des fragments d'ADN insérés déterminées par analyse électrophorétique de leurs fragments de restriction BamHI. L'on peut ainsi déterminer que deux phages appelés  $\lambda$ lacHBs-1 et  $\lambda$ lacHBs-2 correspondant au plasmide pXbaHBs-1 et pXbaHBs-2 contenaient un fragment HBs correctement orienté.

La figure 2a est une carte schématique du vecteur  $\lambda$ plac 5-1 avant sa modification par le fragment HBs-1, issu du pXbaHBs-1.

La figure 2c est une carte schématique d'une partie de ce même vecteur montrant la modification introduite dans son gène Z par insertion dans son site EcoRI du susdit fragment HBs-1.

La figure 2c schématise les structures du polypeptidehybride obtenu comme résultat de l'expression du vecteur modifié de la figure 2b.

L'expression a été obtenue par transfection d'une souche de bactérie de E. coli, notamment d'une souche Hfr $\Delta$ lacX74.

Des souches de E. coli, notamment une souche de E. coli Hfr $\Delta$ lacX74 ont été transformées par plac 5-1,  $\lambda$ lacHBs-1 et  $\lambda$ lacHBs-2 respectivement. Après culture les cellules sont lysées et les lysats obtenus analysés par



électrophorèse sur gel de polyacrylamide SDS (24) et les protéines sont détectées par coloration au bleu de coomassie. On constate la présence d'une bande plus intense parmi les produits d'expression de  $\lambda$ plac 5-1 au niveau de la position correspondante, pour un témoin, à celle de la  $\beta$ -galactosidase (poids moléculaire de 116 248) et d'une bande distincte parmi les produits d'expression de  $\lambda$ lacHBs-1 (non présent parmi les produits d'expression de  $\lambda$ lacHBs-2) correspondant à une nouvelle protéine ayant un poids moléculaire de l'ordre de 135 000-141 000.

Les protéines synthétisées par les bactéries transfectées tant par  $\lambda$ lacHBs-1 que par  $\lambda$ plac 5-1 sont marquées par la (<sup>35</sup>S) méthionine. La mise en présence de ces protéines avec un sérum anti-HBsAg et la réalisation d'un autoradiogramme du gel de polyacrylamide SDS révèlent la présence parmi les produits d'expression du seul  $\lambda$ lacHBs-1 d'une bande à laquelle ne correspond pas de bande équivalente parmi les produits d'expression des autres vecteurs. Cette bande disparaît de façon spécifique lorsque l'immunoprécipitation est réalisée en présence de HBsAg non marqué. On observe encore la même bande parmi les produits d'expression  $\lambda$ lacHBs-1, lorsque l'immunoprécipitation est réalisée avec un antisérum à l'égard de la  $\beta$ -galactosidase.

La structure présumée de la partie de protéine hybride obtenue, au niveau de la fusion entre le gène lacZ et le fragment HBs-1 résulte de la figure 2c qui fait apparaître le fragment " $\beta$ -gal", correspondant à la  $\beta$ -galactosidase (1 005 acides aminés), le fragment HBsAg (192 acides aminés), ces fragments étant séparés par un acide aminé prolyne, correspondant à une partie de "EcoRI linker" contenu dans le vecteur  $\lambda$ lacHBs-1.

C - PROCEDE DE FABRICATION D'UNE MOLECULE IMMUNOGENE METTANT



EN OEUVRE LE VECTEUR SELON L'INVENTION

L'invention peut par conséquent permettre la production d'une protéine de masse moléculaire plus faible que les polypeptides I ou II susvisés, douée des mêmes propriétés immunogènes.

Les résultats démontrent que E. coli, ou tout autre micro-organisme approprié, tel qu'une bactérie ou une culture de cellule eucaryote, peut être infecté par le  $\lambda$ lacHBs-1 et synthétiser une protéine ayant un poids moléculaire de l'ordre de 138 000 et possédant les déterminants antigéniques à la fois de HBsAg et de la  $\beta$ -galactosidase. Cette molécule est représentative des polypeptides hybrides qui peuvent être obtenus par le procédé selon l'invention, dans lequel HBsAg est relié à une protéine support (résultant de la substitution partielle ou totale du fragment  $\beta$ -galactosidase), ces hybrides possédant néanmoins les propriétés antigéniques de HBsAg. Ces nouvelles molécules sont utiles pour la production de vaccins actifs contre l'hépatite virale B.

Comme il va de soi, et comme il résulte d'ailleurs déjà de ce qui précède, l'invention ne se limite nullement à ceux de ses modes d'application et de réalisation qui ont été plus spécialement envisagés ; elle embrasse, au contraire, toutes les variantes.

En annexe de cette description figure la bibliographie, et en particulier les références qui ont été citées dans le cadre de la présente description.



# REFERENCES

- 1 - Blumberg, B. S. (1977) Science 197, 17-25
- 2 - Dane D. S., Cameron C. H., and Briggs M. (1970)  
5 Lancet 1, 695-698
- 3 - WHO Technical Report series, number 602 (1976)
- 4 - Summers J., O'Connell A., and Millman I. (1975) Proc.  
Nat. Acad. Sci. USA 72, 4 597-4 601
- 5 - Hruska J. F., Clayton D. A., Rubenstein J. L. R. and  
10 Robinson W. S. (1977) J. Virol. 21, 666-682
- 6 - Landers T. A., Greenberg H. B. and Robinson W. S.  
(1977) J. Virol. 23, 368-376
- 7 - Charnay P., Pourcel C., Louise A., Fritsch A. and  
Tiollais P. (1979) Proc. Nat. Acad. Sci. USA, 76  
15 2 222-2 226
- 8 - Dreesman G. R., Hollinger F. B., Surians J. R.,  
Fujioka R. B., Brunschwig J. P. and Melnick J. L.  
(1972) J. Virol. 10, 469-476
- 9 - Gerin J. L. (1974) in Mechanisms of virus disease  
20 Ed. W. S. Robinson, C. R. Fox pp 215-24 Menlo Park :  
W. A. Benjamin
- 10 - Dreesman G. R., Chairez R., Suarez M., Hollinger F. B.  
Courtney R. J. and Melnick J. L. (1975) J. Virol. 16,  
508-515
- 25 11 - Shih J. W. and Gerin J. L. (1977) J. Virol. 21,  
1 219-1 222
- 12 - Peterson D. L., Roberts I. M. and Vyas G. N. (1977)  
Proc. Nat. Acad. Sci. USA 74, 1 530-1 534
- 13 - Peterson D. L., Chien D. Y., Vyas G. N., Nitecki D.  
30 and Bond H. (1978) in Viral Hepatitis, Ed. G. Vyas,  
S. Cohen and R. Schmid, The Franklin Institute Press,  
Philadelphia, 569-573
- 14 - Fritsch A., Pourcel C., Charnay P. and Tiollais P.  
(1978) C. R. Acad. Sc. Paris 287, 1 453-1 456
- 35 15 - Burrell C. J., Mackay P., Greenaway P. J.,



- Hofschneider P. H. and Murray K. (1979) Nature 279  
43-47.
- 16 - Tiollais P., Perricaudet M., Pettersson U. and  
Philipson L. (1976) Gène 1, 49-63
- 5 17 - Hérissé J., Courtois G., and Galibert F. (1978)  
Gène 4, 279-294
- 18 - Kroeker W. D. and Laskowski M. S. R. (1977) Anal.  
Biochem. 79, 63-72
- 19 - Maxam A. M. and Gilbert W. (1977) Proc. Nat. Acad.  
10 Sci. USA 74, 560-564
- 20 - Sanger F., Nicklen S. and Coulson A. R. (1977) Proc.  
Nat. Acad. Sci. USA 74, 5 463-5 467
- 21 - Maat J. and Smith A. J. W. (1978) Nucleic Acid. Res.  
5, 4 537-4 545
- 15 22 - Berget S. M., Moore C. and Sharp P. A. (1977) Proc.  
Nat. acad. Sci. USA 74, 3 171-3 175
- 23 - Chow L. T., Gelinas R. E., Broker T. R., and Roberts J.  
(1977), Cell 12, 1-8
- 24 - Shiraishi H., Kohama T., Shirachi R. and Ishida N.  
20 (1977) J. Gen. Virol. 36, 207-210
- 25 - Struck D. K., Lennarz W. J., and Brew K. (1978) J.  
Biol. Chem. 253, 5 786-5 794
- 26 - Reddy V. B., Thimmappaya B., Dhar R., Subramanian K. N.  
Zain B. S., Pan J., Celma C. L. and Weissman S. M.  
25 (1978) Science 200, 494-502
- 27 - Fiers W., Contreras R., Hargeman G., Rogiers R., Van  
de Voorde A., Van Henverswyn H., Van Herreweghe J.,  
Volchaerts G. and Ysebaert M. (1978) Nature 273,  
113-117
- 30 28 - Sanger F., Air G. M., Barrell B. G., Brown N. L., Coulson A. R.,  
Fiddes J. C., Hutchison III C. A., Slocombe P. M. and Smith M.  
(1977) Nature 265, 687-691
- 29 - Barrell B. G., Shaw D. C., Walker J. E., Northrop F. D.,  
Godson G. N. and Fiddes J. C. (1978) Biochem. Soc.  
35 Trans. 6, 63-67

- 30 - Szmunness W., Am. J. Path. 81, 629-649 (1975)
- 31 - Shinsky J. J., Siddiqui A., Robinson W. S. and Cohen S. N., Nature 279, 346-348 (1979)
- 32 - Valenzuela P. et al., Nature 280, 815-819 (1979)
- 5 33 - Charnay P. et al., Nucl. Acids Res. 7, 335-346 (1979)
- 34 - Galibert F., Mandart E., Fitoussi F., Tiollais P. and Charnay P., Nature 281, 646-650 (1979)
- 35 - Pasek M. et al., Nature 282, 575-579 (1979)
- 36 - Vyas G. N., Williams E. W., Klaus G. G. B. and Bond. H. J. Immunol. 108, 1 114-1 118 (1972)
- 10 37 - Hollinger F. B., Dreesman G. R., Sanchez Y., Cabral G. and Melnick J. L., in Viral Hepatitis (Ed. Vyas G. N. Cohen S. N. and Schmid R.) Franklin Institute, Philadelphia, (1978)
- 15 38 - Purcell R. H., and Gerin J. L., Am. J. Med. Sci. 270, 395-399 (1975)
- 39 - Hilleman M. R. et al., Am. J. Med. Sci. 270, 401-404 (1975)
- 40 - Maupas P., Coursaget P., Goudeau A. and Drucker J., Lancet 1, 1 367-1 370 (1976)
- 20 41 - Emtage J. S. et al., Nature 283, 171-174 (1980)
- 42 - Itakura K. et al., Science 198, 1 056-1 063 (1977)
- 43 - Goeddel D. V. et al., Proc. Nat. Acad. Sci. USA 76, 106-110 (1979)
- 25 44 - Pourcel C., Marchal C., Louise A., Fritsch A. and Tiollais P., Molec. Gén. Genet. 170, 161-169 (1979)
- 45 - Bolivar F. et al., Gene 2, 95-113 (1977)
- 46 - Fowler A. V. and Zabin I., Proc. Nat. Acad. Sci. USA 74, 1 507-1 510 (1977)
- 30 47 - Laemmli U. K., Nature 227, 680-685 (1970)
- 48 - Burgess R. R., J. Biol. Chem. 244, 6 168-6 176 (1969)
- 49 - Bonner W. M. and Laskey R. A., Eur. J. Biochem. 46, 83-88 (1974)
- 50 - Laskey R. A. and Mills A. D., Eur. J. Biochem. 56, 335-341 (1975)
- 35



- 51 - Iwakura Y., Ito K. and Ishihama A., Molec. Gen. Genet.  
133, 1-23 (1974)
- 52 - Talwai G. P., et al., Proc. Nat. Acad. Sci. USA 73,  
218-222 (1976)



REVENDICATIONS

1. Acide nucléique caractérisé en ce qu'il comporte au plus de l'ordre de 1 000-1 100 nucléotides et qu'il est apte à coder une séquence peptidique immunogène, 5 elle-même apte à induire in vivo la production d'anticorps actifs à l'égard du virus de l'hépatite B.

2. Acide nucléique selon la revendication 1, caractérisé en ce qu'il comporte une séquence nucléotidique capable de coder la séquence peptidique représentée dans 10 les figures 3A, 3B, 3C ou une séquence peptidique analogue douée de propriétés immunogènes équivalentes.

3. Acide nucléique selon la revendication 2, caractérisé en ce qu'il comporte une séquence nucléotidique représentée aux figures 4A, 4B, capable de coder la séquen- 15 ce peptidique représentée dans la figure 5 ou une séquence peptidique analogue douée de propriétés immunogènes équivalentes.

4. Acide nucléique selon les revendications 1 à 3, caractérisé en ce qu'il comporte une séquence nucléo- 20 tidique capable de coder la séquence peptidique représentée à la figure 6 ou une séquence peptidique analogue douée de propriétés immunogènes équivalentes représentée à la figure 6.

5. Acide nucléique selon l'une quelconque des 25 revendications 1 à 4, caractérisé en ce qu'il contient une séquence capable de coder la séquence peptidique : Alanine (N-terminale) - Glutamine - Glycine - Thréonine - Sérine (C-terminale).

6. Acide nucléique selon l'une quelconque des 30 revendications 1 à 4, caractérisé en ce qu'il contient une séquence capable de coder la séquence peptidique : Thréonine (N-terminale) - Alanine - Glutamine - Glycine - Thréonine - Sérine (C-terminale).

7. Acide nucléique selon l'une quelconque des 35 revendications 1 à 4, caractérisé en ce qu'il contient

FEUILLE DE L'EMPLACEMENT



une séquence capable de coder la séquence peptidique :  
Thréonine (N-terminale) - Thréonine - Alanine - Glutamine -  
Glycine - Thréonine - Sérine (C-terminale).

8. Acide nucléique selon l'une quelconque des  
5 revendications 1 à 7, caractérisé en ce qu'il comporte au  
moins l'un des deux brins mutuellement complémentaires d'une  
séquence d'ADN telle que représentée dans les figures 4A, 4B ou un  
brin susceptible de s'hybrider avec le précédent.

9. Acide nucléique selon la revendication 8,  
10 caractérisé en ce qu'il comporte au moins l'un des deux  
brins mutuellement complémentaires d'une séquence d'ADN  
telle que représentée dans les figures 3A, 3B, 3C ou un brin suscep-  
tible de s'hybrider avec le précédent.

10. Acide nucléique selon l'une quelconque des  
15 revendications 1 à 7, caractérisé en ce qu'il est à double  
brin et en ce qu'il est délimité à l'un de ses bouts par  
une extrémité HincII, HhaI, AvaI, ou EcoRI, et à son autre  
bout par une extrémité AvaIII, HincII, ou HhaI.

11. Acide nucléique selon la revendication 5,  
20 caractérisé en ce qu'il contient la séquence 5' GCT CAA  
GGA ACC TCT 3' ou la séquence complémentaire de la précé-  
dente, ou encore une séquence dans laquelle l'un au moins  
des susdits triplets est remplacé par un autre triplet,  
cependant capable de coder le même amino-acide.

25 12. Acide nucléique selon la revendication 6,  
caractérisé en ce qu'il contient la séquence 5' ACT GCT  
CAA GGA ACC TCT 3', ou la séquence complémentaire de la  
précédente, ou encore une séquence dans laquelle l'un au  
moins des susdits triplets est remplacé par un autre tri-  
30 plet, cependant capable de coder le même amino-acide.

13. Acide nucléique selon la revendication 7,  
caractérisé en ce qu'il contient la séquence 5' ACT ACT  
GCT CAA GGA ACC TCT 3', ou la séquence complémentaire de  
la précédente, ou encore une séquence dans laquelle l'un  
35 au moins des susdits triplets est remplacé par un autre

FEUILLE DE REMPLACEMENT



triplet, cependant capable de coder le même amino-acide.

14. Application de l'ADN selon l'une quelconque des revendications 1 à 13, à la constitution d'un vecteur susceptible d'être exprimé dans un micro-organisme ou dans  
5 une cellule eucaryote en vue de la production d'une protéine comportant une séquence peptidique codée par ledit acide nucléique.

15. Peptide, caractérisé en ce qu'il comporte une séquence peptidique codée par l'acide nucléique selon  
10 l'une quelconque des revendications 1 à 13.

16. Peptide selon la revendication 15, caractérisé par la séquence suivante : Alanine - Glutamine - Glycine - Thréonine - Sérine, dans laquelle l'extrémité alanine est N-terminale, et l'extrémité sérine est C-  
15 terminale.

17. Peptide selon la revendication 15, caractérisé par la séquence suivante : Thréonine - Alanine - Glutamine - Glycine - Thréonine - Sérine, dans laquelle l'extrémité thréonine est N-terminale et l'extrémité séri-  
20 ne est C-terminale.

18. Peptide selon la revendication 15, caractérisé par la séquence suivante : Thréonine - Thréonine - Alanine - Glutamine - Glycine - Thréonine - Sérine, dans laquelle l'extrémité thréonine est N-terminale et l'extré-  
25 mité sérine est C-terminale.

19. Composition pharmaceutique, notamment de vaccin ou de réactif de diagnostic de l'hépatite B, contenant en association avec un véhicule pharmaceutiquement acceptable le polypeptide selon l'une quelconque des reven-  
30 dications 16, 17 et 18, soit seul, soit au préalable fixé sur une molécule porteuse, notamment de type polypeptide ou protéine.

20. Vecteur, notamment du type phage ou plasmide selon l'une quelconque des revendications 1 à 18, con-  
35 tenant au moins une partie de l'opéron lactose, plus



particulièrement le promoteur et le gène Z de cet opéron, ce vecteur étant caractérisé en ce qu'il est modifié pour l'insertion, en phase, dans un site approprié du gène Z, tel que le site EcoRI, de l'un quelconque des fragments 5 d'ADN du brevet principal notamment ceux contenant la plus grande partie du gène S.

21. Vecteur selon la revendication 20, caractérisé en ce qu'au moins une partie du fragment d'ADN codant pour la plus grande partie de la  $\beta$ -galactosidase est remplacée par un fragment d'ADN apte à coder pour toute autre molécule porteuse non immunogène, ou dont les éventuelles propriétés immunologiques, si celles-ci existent, n'interfèrent pas avec celles de la partie peptidique présentant les propriétés immunologiques de HBsAg.

15 22. Vecteur selon la revendication 21, caractérisé en ce que le susdit fragment d'ADN de remplacement s'étend dans la direction de la lecture à partir du site HhaI antérieurement contenu dans le fragment d'ADN codant pour la plus grande partie de la  $\beta$ -galactosidase.

20 23. Vecteur selon l'une quelconque des revendications 20 à 22, caractérisé en ce que le vecteur est dérivé du plasmide pBR322.

24. Vecteur selon l'une quelconque des revendications 20 à 23, caractérisé en ce que le fragment inséré 25 susdit est délimité par des extrémités EcoRI et qu'il contient un fragment originaire d'ADN provenant du virus de l'hépatite B, qui, dans celui-ci correspond au fragment délimité par des sites HhaI et XbaI et contenant lui-même la majeure partie du gène S.

30 25. Vecteur selon l'une quelconque des revendications 20 à 24, caractérisé en ce qu'il comprend une séquence nucléotidique d'environ 1 200 nucléotides provenant d'un virus de l'hépatite B et en ce qu'il se multiplie dans un micro-organisme ou dans une cellule eucaryote.

35 26. La protéine hybride susceptible d'être





codée par le fragment d'ADN inséré dans le susdit vecteur.

27. Protéine hybride selon la revendication 26, contenant une séquence polypeptidique présentant les propriétés immunologiques spécifiques de HBsAg, jouxtant une 5 séquence polypeptidique constituée par la majeure partie de la  $\beta$ -galactosidase.

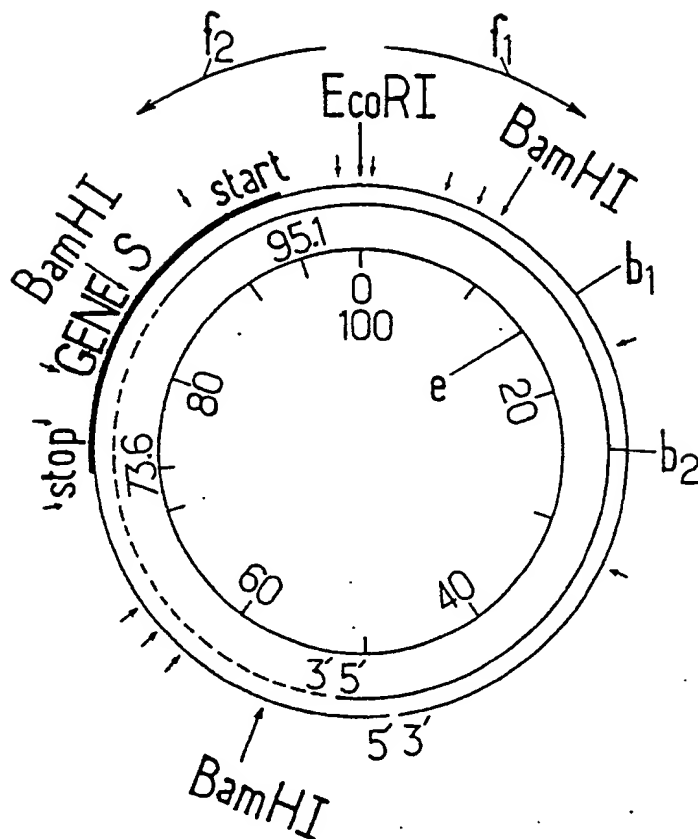
28. Protéine hybride selon la revendication 27, caractérisée en ce que tout ou partie de la partie  $\beta$ -galactosidase est remplacée par toute autre molécule porteuse 10 non immunogène, ou dont les éventuelles propriétés immunologiques, si celles-ci existent, n'interfèrent pas avec celles de la partie peptidique présentant les propriétés immunologiques de HBsAg.

29. Composition de vaccin contenant à titre de 15 principe actif la protéine hybride selon l'une quelconque des revendications 26 à 28.



I/13

Fig.1.



FEUILLE DE REMPLACEMENT



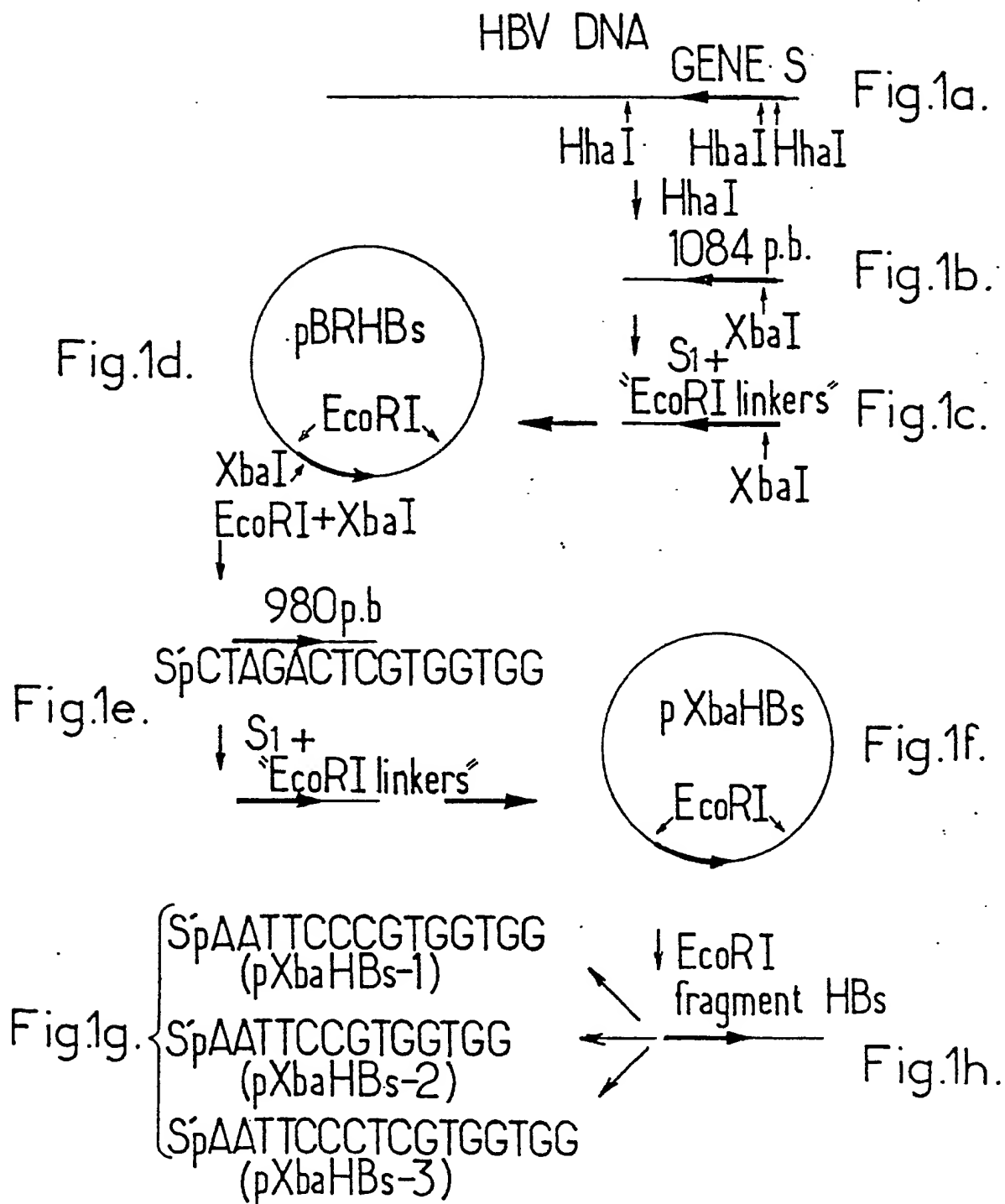


Fig.2A.

Hinc II (2191)

Ava III (2116)		HhaI (1952)			
1930	1940	1950	1960	1970	1980
TCGGCAGAGG	AGCCGAAVAG	GTTCCACGCA	TGCGCTGATG	GCCCATGACC	AAGCCCCAGC
AGCCGTCTCC	TCGGCTTTTC	CAAGGTGCGT	ACGCGACTAC	CGGGTACTGG	TTGGGGTCTG
1990	2000	2010	2020	2030	2040
CAGTGGGGGT	TGCGTCAGCA	AACACTTGGC	ACAGACCTGG	CCGTTGCCGG	GCAACGGGGT
GTCACCCCCA	ACGCAGTCGT	TTGTGAACCG	TGTCTGGACC	GGCAACGGCC	CGTTGCCCCA
2050	2060	2070	2080	2090	2100
MAAGGTTGAG	GTATTGTTTA	CACAGAAAGG	CCTTGTAAGT	TGGCGAGAAA	GTGAAAGCCT
TTTCCVAGTC	CATAACAAAT	GTGTCTTTCC	GGAACATTCA	ACCGCTCTTT	CACTTTCGGA
2110	2120	2130	2140	2150	2160
GCTTAGATTG	AATACATGCA	TACAAAGGCA	TCAACGCAGG	ATAACCACAT	TGTGTAAAG
CGAATCTAAC	TTATGTACGT	ATGTTTCCGT	AGTTGCGTCC	TATTGGTGTA	ACACATTTTC
2170	2180	2190	2200	2210	2220
GGGCAGCAAA	ACCCAAVAGA	CCCACAATTC	GTTGACATAC	TTTCCAATCA	ATAGGCCTGT
CCCGTCGTTT	TGGGTTTTCT	GGGTGTTAAG	CAACTGTATG	AAAGGTTAGT	TATCCGGACA
2230	2240	2250	2260	2270	2280
TATAGGAAAG	TTTTCTAAV	CATTCTTTGA	TTTTTTGTAT	GATGTGTTCT	TGTGGCAAGG
ATTATCCTTC	AAVAGATTTT	GAAAGAACT	AAAAAACATA	CTACACAAGA	ACACCGTTCC
2290	2300	2310	2320	2330	2340
ACCCATVACA	TCCVMTGACA	TAVCCCATAA	AATTTAGAGA	GTAACCCCAT	CTCTTTGTTT
TGGGTATTGT	AGGTTACTGT	AITGGGTATT	TTAAATCTCT	CATTGGGGTA	GAGAAACAAA
2350	2360	2370	2380	2390	2400
TGTTAGGGTT	TAAVMTGATA	CCCAAAGACA	AAVGAATTT	GGTAACAGCG	GTAAAAAGGG
ACVATCCCAA	ATTACATAT	GGGTTTCTGT	TTTCTTTTAA	CCATTGTCGC	CATTTTTCCC
2410	2420	2430	2440	2450	2460
ACTCAGATG	CTGTACAGAC	TTGGCCCCCA	ATACCACATC	ATCCATATV	CTGAAVGCCA
TGAGTTCTAC	GACATGTCTG	AACCGGGGGT	TATGGTGTAG	TAGGTATATT	GACTTTCGGT
2470	2480	2490	2500	2510	2520
AAVAGTGSGG	GAAVGCCTA	CGAACCCTG	AACAAATGGC	ACTAGTAAAC	TGAGCCAGGA
TTGTCACCCC	CTTTCGGGAT	CGTTGGTGAC	TTGTTTACCG	TGATCATTTG	ACTCGGTCCT

t S''

FEUILLE DE REMPLACEMENT



IV/13

Fig.2B.

2530	2540	2550	2560	2570	2580
GAAACGGGCT	GAGGCCCACT	CCCATAGGAA	TTTTCCGAAA	GCCCAGGATG	ATGGGATGGG
CTTTGCCCGA	CTCCGGGTGA	GGGTATCCTT	AAAAGGCTTT	CGGGTCCTAC	TACCCTACCC
2590	2600	2610	2620	2630	2640
AATACAGGTG	CAATTTCCGT	CCGAAGGTTT	GGTACAGCAA	CAGGAGGGAT	ACATAGAGGT
TTATGTCCAC	GTAAAGGCA	GGCTTCCAAA	CCATGTCGTT	GTCCTCCCTA	TGTATCTCCA
2650	2660	2670	2680	2690	2700
TCCTTGAGCA	GTAGTCATGC	AGGTCCGGCA	TGGTCCCGTG	CTGGTTGTTG	AGGATCCTGG
AGGAACTCGT	CATCAGTACG	TCCAGGCCGT	ACCAGGGCAC	GACCAACAAC	TCCTAGGACC
2710	2720	2730	2740	2750	2760
AATTAGAGGA	CAAACGGGCA	ACATACCTTG	ATAGTCCAGA	AGAACCAACA	AGAAGATGAG
TTAATCTCCT	GTTTGCCCGT	TGTATGGAAC	TATCAGGTCT	TCTTGTTGT	TCTTCTACTC
2770	2780	2790	2800	2810	2820
GCATAGCAGC	AGGATGAAGA	GGAAGATGAT	AAAACGCCGC	AGACACATCC	AGCGATAACC
CGTATCGTCG	TCCTACTTCT	CCTTCTACTA	TTTTGCGGCG	TCTGTGTAGG	TCGCTATTGG
2830	2840	2850	2860	2870	2880
AGGACAAGTT	GGAGGACAAG	AGGTTGGTGA	GTGATTGGAG	GTTGGGGACT	GCGAATTTTG
TCCTGTTCAA	CCTCCTGTTT	TCCAACCACT	CACTAACCTC	CAACCCCTGA	CGCTTAAAC
2890	2900	2910	2920	2930	2940
GCCAAGACAC	ACGGTAGTTC	CCCCTAGAAA	ATTGAGAGAA	GTCCACCACG	AGTCTAGACT
CGGTTCTGTG	TGCCATCAAG	GGGGATCTTT	TAACCTCTTT	CAGGTGGTGC	TCAGATCTGA
2950	2960	2970	2980	2990	3000
CTGCGGTATT	GTGAGGATTC	TTGTCAACAA	GAAAAACCCC	GCCTGTAACA	CGAGAAGGGG
GACGCCATAA	CACTCCTAAG	AACAGTTGTT	CTTTTGGGG	CGGACATTGT	GCTCTTCCCC
3010	3020	3030	3040	3050	3060
TCCTAGGAAT	CCTGATGTGA	TGTTCTCCAT	GTTCAAGCGA	GGGTCCCCAA	TCCTCGAGAA
AGGATCCTTA	GGACTACACT	ACAAGAGGTA	CAAGTCGCGT	CCCAGGGGTT	AGGAGCTCTT
3070	3080	3090	3100	3110	3120
GATTGACGAT	AAGGGAGAGG	CAGTAGTCAG	AACAGGGTTT	ACTGTTCTTG	AACTGGAGCC
CTAACTGCTA	TTCCCTCTCC	GTCATCAGTC	TTGTCCCAAA	TGACAAGGAC	TTGACCTCGG

Hinc II(2963)

p<sup>5</sup>S'

Hha I(3036)

Ava I(3053)

FEUILLE DE REMPLACEMENT



V/13

## Fig.2C.

3130	3140	3150	3160	3170	3180
ACCAGCAGGG	AAATACAGGC	CTCTCACTCT	GGGATCTTGC	AGAGTTTGGT	GGAAGGTTGT
TGGTCGTCCC	TTTATGTCCG	GAGAGTGAGA	CCCTAGAACG	TCTCAAACCA	CCTTCCAACA

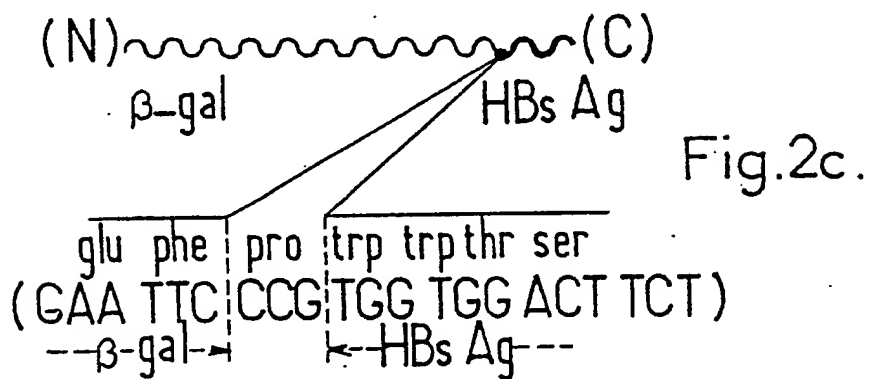
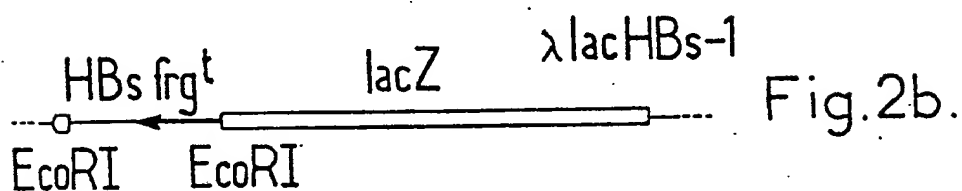
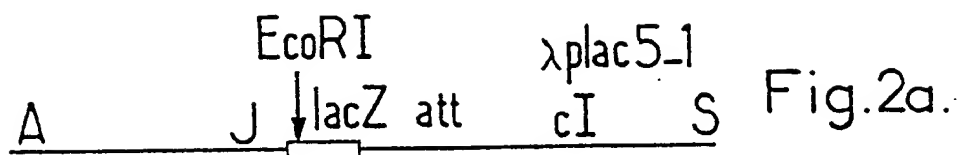
GG L 3'  
CC S 5'

f<sub>2</sub>

EcoRI (1/3182)

FEUILLE DE REMPLACEMENT





FEUILLE DE REMPLACEMENT



VII 13

 $f_2$   

Fig. 3A.

										30
3'	TAC	CTC	TTG	TAG	TGT	AGT	CCT	AAG	GAT	CCT
5'	ATG	GAG	AAC	ATC	ACA	TCA	GGA	TTC	CTA	GGA
	MET	GLU	ASN	ILE	THR	SER	GLY	PHE	LEU	GLY
										60
	GCG	GAA	GAG	CAC	AAT	GTC	CGC	CCC	AAA	AAG
	CCC	CTT	CTC	GTG	TTA	CAG	GCG	GGG	TTT	TTC
	PRO	LEU	LEU	VAL	LEU	GLN	ALA	GLY	PHE	PHE
										90
	AAC	AAC	TGT	TCT	TAG	GAG	TGT	TAT	GGC	GTC
	TTG	TTG	ACA	AGA	ATC	CTC	ACA	ATA	CCG	CAG
	LEU	LEU	THR	ARG	ILE	LEU	THR	ILE	PRO	GLN
										120
	TCA	GAT	CTG	AGC	ACC	ACC	TGA	AGA	GAG	TTA
	AGT	CTA	GAC	TCG	TGG	TGG	ACT	TCT	CTC	AAT
	SER	LEU	ASP	SER	TRP	TRP	THR	SER	LEU	ASN
										150
	AAA	GAT	CCC	CCT	TGA	TGG	CAC	ACA	GAA	CCG
	TTT	CTA	GGG	GGA	ACT	ACC	GTG	TGT	CTT	GGC
	PHE	LEU	GLY	GLY	THR	THR	VAL	CYS	LEU	GLY
										180
	GTT	TTA	AGC	GTC	AGG	GGT	TGG	AGG	TTA	GTG
	CAA	AAT	TCG	CAG	TCC	CCA	ACC	TCC	AAT	CAC
	GLN	ASN	SER	GLN	SER	PRO	THR	SER	ASN	HIS
										210
	AGT	GGT	TGG	AGA	ACA	GGA	GGT	TGA	ACA	GGA
	TCA	CCA	ACC	TCT	TGT	CCT	CCA	ACT	TGT	CCT
	SER	PRO	THR	SER	CYS	PRO	PRO	THR	CYS	PRO
										240
	CCA	ATA	GCG	ACC	TAC	ACA	GAC	GCC	GCA	AAA
	GGT	TAT	CGC	TGG	ATG	TGT	CTG	CCG	CGT	TTT
	GLY	TYR	ARG	TRP	MET	CYS	LEU	ARG	ARG	PHE

FEUILLE DE REMPLACEMENT





## Fig.3B. VIII/13

270  
TAG TAG AAG GAG AAG TAG GAC GAC GAT ACG  
ATC ATC TTC CTC TTC ATC CTG CTG CTA TGC  
ILE ILE PHE LEU PHE ILE LEU LEU LEU CYS

300  
GAG TAG AAG AAC AAC CAA GAA GAC CTG ATA  
CTC ATC TTC TTG TTG GTT CTT CTG GAC TAT  
LEU ILE PHE LEU LEU VAL LEU LEU ASP TYR

330  
GTT CCA TAC AAC GGG CAA ACA GGA GAT TAA  
CAA GGT ATG TTG CCC GTT TGT CCT CTA ATT  
GLN GLY MET LEU PRO VAL CYS PRO LEU ILE

Site Bam HI 360  
GGT CCT AGG AGT TGT TGG TCG TGC CCT GGT  
CCA GGA TCC TCA ACA ACC AGC ACG GGA CCA  
PRO GLY SER SER THR THR SER THR GLY PRO

390  
ACG GCC TGG ACG TAC TGA TGA CGA GTT CCT  
TGC CGG ACC TGC ATG ACT ACT GCT CAA GGA  
CYS ARG THR CYS MET THR THR ALA GLN GLY

420  
TGG AGA TAC ATA GGG AGG ACA ACG ACA TGG  
ACC TCT ATG TAT CCC TCC TGT TGC TGT ACC  
THR SER MET TYR PRO SER CYS CYS CYS THR

450  
TTT GGA AGC CTG CCT TTA ACG TGG ACA TAA  
AAA CCT TCG GAC GGA AAT TGC ACC TGT ATT  
LYS PRO SER ASP GLY ASN CYS THR CYS ILE

480  
GGG TAG GGT AGT AGG ACC CGA AAG CCT TTT  
CCC ATC CCA TCA TCC TGG GCT TTC GGA AAA  
PRO ILE PRO SER SER TRP ALA PHE GLY LYS

FEUILLE DE REMPLACEMENT



IX/13

## Fig.3C.

510

AAG	GAT	ACC	CTC	ACC	CGG	AGT	CGG	GCA	AAG
TTC	CTA	TGG	GAG	TGG	GCC	TCA	GCC	CGT	TTC

Site Hae III

PHE	LEU	TRP	GLU	TRP	ALA	SER	ALA	ARG	PHE
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

540

AGG	ACC	GAG	TCA	AAT	GAT	CAC	GGT	AAA	CAA
TCC	TGG	CTC	AGT	TTA	CTA	GTG	CCA	TTT	GTT
SER	TRP	LEU	SER	LEU	LEU	VAL	PRO	PHE	VAL

570

GTC	ACC	AAG	CAT	CCC	GAA	AGG	GGG	TGA	CAA
CAG	TGG	TTC	GTA	GGG	CTT	TCC	CCC	ACT	GTT
GLN	TRP	PHE	VAL	GLY	LEU	SER	PRO	THR	VAL

600

ACC	GAA	AGT	CAA	TAT	ACC	TAC	TAC	ACC	ATA
TGG	CTT	TCA	GTT	ATA	TGG	ATG	ATG	TGG	TAT
TRP	LEU	SER	VAL	ILE	TRP	MET	MET	TRP	TYR

630

ACC	CCC	GGT	TCA	GAC	ATG	TCG	TAG	AAC	TCA
TGG	GGG	CCA	AGT	CTG	TAC	AGC	ATC	TTG	AGT
TRP	GLY	PRO	SER	LEU	TYR	SER	ILE	LEU	SER

690

GGG	AAA	AAT	GGC	GAC	AAT	GGT	TAA	AAG	AAA
CCC	TTT	TTA	CCG	CTG	TTA	CCA	ATT	TTC	TTT
PRO	PHE	LEU	PRO	LEU	LEU	PRO	ILE	PHE	PHE

ACA	GAA	ACC	CAT	ATG	TAA	ATT	5'
TGT	CTT	TGG	GTA	TAC	ATT	TAA	3'
CYS	LEU	TRP	VAL	TYR	ILE	STOP	

FEUILLE DE REMPLACEMENT



X/13

Fig.4A.

	2191	2200	2210	2220
	5' .TTGACATAC	TTTCCAATCA	ATAGGCCTGT	
	3' AACTGTATG	AAAGGTTAGT	TATCCGGACA	
2230	2240	2250	2260	2270
TAATAGGANG	TTTTCTAAAA	CATTCTTTGA	TTTTTTGTAT	GATGTGTTCT
ATTATCCTTC	AAAAGATTTT	GTAAGAACT	AAAAAACATA	CTACACAAGA
				2280
2290	2300	2310	2320	2330
ACCCATAACA	TCCAATGACA	TAACCCATAA	AATTTAGAGA	GTAACCCCAT
TGGGTATTGT	AGGTTACTGT	ATTGGGTATT	TAAATCTCT	CATTGGGGTA
				2340
2350	2360	2370	2380	2390
TGTTAGGGTT	TAAATGTATA	CCCAAAGACA	AAAGAAAATT	GGTAACAGCG
ACAATCCCAA	ATTTACATAT	GGGTTTCTGT	TTTCTTTTAA	CCATTGTCGC
				2400
2410	2420	2430	2440	2450
ACTCAAGATG	CTGTACAGAC	TTGGCCCCCA	ATACCACATC	ATCCATATAA
TGAGTTCTAC	GACATGTCTG	AACCGGGGGT	TATGGTGTAG	TAGGTATATT
				2460
2470	2480	2490	2500	2510
AACAGTGGGG	GAAAGCCCTA	CGAACCACTG	AACAAATGGC	ACTAGTAAAC
TTGTCACCCC	CTTTGCGGAT	GCTTGGTGAC	TTGTTTACCG	TGATCATTG
				2520
2530	2540	2550	2560	2570
GAAACGGGCT	GAGGCCCACT	CCCATAGGAA	TTTTCCGAAA	GCCCAGGATG
CTTTGCCCGA	CTCCGGGTGA	GGGTATCCTT	AAAAGGCTTT	CGGGTCCTAC
				2580
2590	2600	2610	2620	2630
AATACAGGTG	CAATTTCCGT	CGAAGGTTT	GGTACAGCAA	CAGGAGGGAT
TTATGTCCAC	GTTAAAGGCA	GGCTTCCAAA	CCATGTCGTT	GTCCTCCCTA
				2640
2650	2660	2670	2680	2690
TCCTTGAGCA	GTAGTCATGC	AGGTCCGGCA	TGGTCCCGTG	CTGGTTGTTG
AGGAACCTCGT	CATCAGTACG	TCCAGGCCGT	ACCAGGGCAC	GACCAACAAC
				2700
2710	2720	2730	2740	2750
AATTAGAGGA	CAAACGGGCA	ACATACCTTG	ATAGTCCAGA	AGAACCAACA
TTAATCTCCT	GTTTGCCCGT	TGTATGGAAC	TATCAGGTCT	TCTTGTTGT
				2760
2770	2780	2790	2800	2810
GCATAGCAGC	AGGATGAAGA	GGAAGATGAT	AAAACGCCGC	AGACACATCC
CGTATCGTCG	TCCTACTTCT	CCTTCTACTA	TTTTGCGGCG	TCTGTGTAGG
				2820
				AGCGATAACC
				TCGCTATTGG

FEUILLE DE REMPLACEMENT



XI/13

Fig.4B.

2830	2840	2850	2860	2870	2880
AGGACAAGTT	GGAGGACAAG	AGGTTGGTGA	GTGATTGGAG	GTTGGGGACT	GCGAATTTTG
TCCTGTTCAA	CCTCCTGTTT	TCCAACCACT	CACTAACCTC	CAACCCCTGA	CGCTTAAAC
2890	2900	2910	2920	2930	2940
GCCAAGACAC	ACGGTAGTTC	CCCCTAGAAA	ATTGAGAGAA	GTCCACCACG	AGTCTAGACT
CGGTTCTGTG	TGCCATCAAG	GGGGATCTTT	TAACTCTCTT	CAGGTGGTGC	TCAGATCTGA
2950	2960				
CTGCGGTATT	GTGAGGATTC	TTG L 3'			
GACGCCATAA	CACTCCTAAG	AAC S 5'			

← f<sub>2</sub>

FEUILLE DE REMPLACEMENT



XII/13

Fig.5.

24										30				
ARG	ILE	LEU	THR	ILE	PRO	GLN	SER	LEU	ASP	SER	TRP			
40										50				
TRP	THR	SER	LEU	ASN	PHE	LEU	GLY	GLY	THR	THR	VAL	CYS	LEU	GLY
60														
GLN	ASN	SER	GLN	SER	PRO	THR	SER	ASN	HIS	SER	PRO	THR	SER	CYS
70										80				
PRO	PRO	THR	CYS	PRO	GLY	TYR	ARG	TRP	MET	CYS	LEU	ARG	ARG	PHE
90														
ILE	ILE	PHE	LEU	PHE	ILE	LEU	LEU	LEU	CYS	LEU	ILE	PHE	LEU	LEU
100										110				
VAL	LEU	LEU	ASP	TYR	GLN	GLY	MET	LEU	PRO	VAL	CYS	PRO	LEU	ILE
120														
PRO	GLY	SER	SER	THR	THR	SER	THR	GLY	PRO	CYS	ARG	THR	CYS	MET
130										140				
THR	THR	ALA	GLN	GLY	THR	SER	MET	TYR	PRO	SER	CYS	CYS	CYS	THR
150														
LYS	PRO	SER	ASP	GLY	ASN	CYS	THR	CYS	ILE	PRO	ILE	PRO	SER	SER
160										170				
TRP	ALA	PHE	GLY	LYS	PHE	LEU	TRP	GLU	TRP	ALA	SER	ALA	ARG	PHE
180														
SER	TRP	LEU	SER	LEU	LEU	VAL	PRO	PHE	VAL	GLN	TRP	PHE	VAL	GLY
190										200				
LEU	SER	PRO	THR	VAL	TRP	LEU	SER	VAL	ILE	TRP	MET	MET	TRP	TYR
210														
TRP	GLY	PRO	SER	LEU	TYR	SER	ILE	LEU	SER	PRO	PHE	LEU	PRO	LEU
220										226				
LEU	PRO	ILE	PHE	PHE	CYS	LEU	TRP	VAL	TYR	ILE				

# FEUILLE DE REMPLACEMENT



XIII/13

[illegible]

# FEUILLE DE REMPLACEMENT



# RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale N°PCT/FR 80/00133

<b>I. CLASSEMENT DE L'INVENTION</b> (si plusieurs symboles de classification sont applicables, les indiquer tous) <sup>1</sup>		
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB Int.Cl. <sup>3</sup> C 12 N 15/00; C 07 H 21/04; C 12 P 21/02; C 07 C 103/52; A 01 K 39/29; G 01 N 33/54		
<b>II. DOMAINES SUR LESQUELS LA RECHERCHE A PORTÉ</b>		
Documentation minimale consultée <sup>4</sup>		
Système de classification	Symboles de classification	
Int.Cl. <sup>3</sup>	C 12 N 15/00; C 12 P 21/00; C 12 P 21/02; C 07 H 21/04	
Documentation consultée autre que la documentation minimale dans la mesure où de tels documents font partie des domaines sur lesquels la recherche a porté <sup>5</sup>		
<b>III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS</b> <sup>14</sup>		
Catégorie *	Identification des documents cités, <sup>16</sup> avec indication, si nécessaire, des passages pertinents <sup>17</sup>	N° des revendications visées <sup>18</sup>
X	Nature, volume 279, publié le 3 mai 1979, (USA) C.J. Burrell et al. "Expression in Escherichia coli of hepatitis B virus DNA sequences cloned in plasmid pBR 322", pages 43 à 47; voir figures 1,2,3; page 45, colonne 1, paragraphe 1 à page 47, fin cité dans la demande	1 à 4, 8 à 10,14, 15
X	Proc. Natl. Acad. Sci, volume 76, no. 5, publié en mai 1979, (USA) P. Charnay et al. "Cloning in Escherichia coli and physical structure of hepatitis B virion DNA", pages 2222-2226, voir figures 2,3,5; page 2222: "Materials and Methods" cité dans la demande	1 à 4, 8 à 10,14
X	Nature, volume 279, publié le 24 mai 1979, J.J. Sninsky et al. "Cloning and endonuclease mapping of the hepatitis B viral genome", pages 346-348; voir l'article en entier cité dans la demande	1 à 4,8 à 10,14
* Catégories spéciales de documents cités: <sup>14</sup> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>« A » document définissant l'état général de la technique</p> <p>« E » document antérieur, mais publié à la date de dépôt international ou après cette date</p> <p>« L » document cité pour raison spéciale autre que celles qui sont mentionnées dans les autres catégories</p> <p>« O » document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens</p> </div> <div style="width: 45%;"> <p>« P » document publié avant la date de dépôt international mais à la date de priorité revendiquée ou après celle-ci</p> <p>« T » document ultérieur publié à la date de dépôt international ou à la date de priorité, ou après, et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention</p> <p>« X » document particulièrement pertinent</p> </div> </div>		
<b>IV. CERTIFICATION</b>		
Date à laquelle la recherche internationale a été effectivement achevée <sup>19</sup>	Date d'expédition du présent rapport de recherche internationale <sup>20</sup>	
26 novembre 1980	19 décembre 1980	
Administration chargée de la recherche internationale <sup>1</sup>	Signature du fonctionnaire autorisé <sup>20</sup>	
Office Européen des Brevets	G.L.M. Kruidenberg	

III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS <sup>14</sup> (SUITE DES RENSEIGNEMENTS INDICUÉS SUR LA DEUXIÈME FEUILLE)		
Catégorie <sup>15</sup>	Identification des documents cités, <sup>16</sup> avec indication, si nécessaire des passages pertinents <sup>17</sup>	N° des revendications visées <sup>18</sup>
X	Chemical Abstracts, volume 86, no. 25, publié le 20 juin 1977 (Columbus Ohio, USA) D.P. Peterson et al. "Partial amino acid sequence of the major component polypeptides of hepatitis B surface antigen", voir page 413, colonne 2, l'abrégé 187.460r Proc. Natl. Acad. Sci. USA 1977, 74(4), pages 1530-34 --	15
	C.R. Acad. Sc., volume 287, publié le 18 décembre 1978, (Paris, FR) A. Fritsch et al. "Clonage du genome du virus de l'hépatite B dans Escherichia coli", pages 1453-1456, voir l'article en entier cité dans la demande --	1
P,X	Nature, volume 286, publié le 28 août 1980, P. Charnay et al. "Biosynthesis of hepatitis B virus surface antigen in Escherichia coli", pages 893-895, voir l'article en entier --	14 à 20, 23 à 29
P,X	Nucleic Acids Research, volume 7, no. 2, publié le 25 septembre 1979, P. Charnay et al. "Localization on the viral genome and nucleotide sequence of the gene coding for the two major polypeptides of the Hepatitis B surface antigen (HBSAg). pages 335 à 346, voir le document en entier cité dans la demande --	1 à 13, 15 à 18
F, II	Nature, volume 281, publié le 25 octobre 1979, F. Galibert et al. "Nucleotide sequence of the hepatitis B virus genome (sub-type ayw) cloned in E.coli", pages 646-650, voir l'article en entier cité dans la demande --	1 à 13
P,X	Nature, volume 282, publié le 6 décembre 1979, M. Pasek et al. "Hepatitis B virus genes and their expression in E.coli", pages 575-579, voir l'article en entier cité dans la demande --	1 à 4, 8, 9, 14, 15
P,X	Nature, volume 280, publié le 30 août 1979, P. Valenzuela et al. "Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen", pages 815-819 cité dans la demande --	1 à 4, 8, 9, 14, 15
	--	./.



**SUITE DES RENSEIGNEMENTS INDICUÉS SUR LA DEUXIÈME FEUILLE**

P,E	EP, A, 13828, publié le 6 août 1980, voir page 11, ligne 1 à page 21, ligne 3; page 27, ligne 4 à page 36, fin; revendications 1 à 4, 6 à 32; figures 1 à 12, Biogen N.V.	1 à 4,8 à 10,14,15, 19,20,23, 25,26,27, 29
P,E	DE, A, 2950995, publié le 24 juillet 1980, voir page 4, lignes 5 à 15; page 6, ligne 16 à page 17, fin; revendications 1 à 12, Institut Pasteur	

**V. OBSERVATIONS LORSQU'IL A ÉTÉ ESTIMÉ QUE CERTAINES REVENDICATIONS NE POUVAIENT PAS FAIRE L'OBJET D'UNE RECHERCHE <sup>10</sup>**

Selon l'article 17.2) a) certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants:

- ☐ Les revendications numéros.....se rapportent à un objet à l'égard duquel la présente administration n'a pas l'obligation de procéder à la recherche, <sup>12</sup> à savoir:
- ☐ Les revendications numéros.....se rapportent à des parties de la demande internationale qui ne remplissent pas les conditions prescrites dans une mesure telle qu'une recherche significative ne peut être effectuée, <sup>13</sup> précisément:

**V. OBSERVATIONS LORSQU'IL Y A ABSENCE D'UNITÉ DE L'INVENTION <sup>11</sup>**

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la présente demande internationale, c'est-à-dire:

- ☐ Comme toutes les taxes additionnelles demandées ont été payées dans les délais, le présent rapport de recherche internationale couvre toutes les revendications de la demande internationale pouvant faire l'objet d'une recherche.
- ☐ Comme seulement une des parties taxes additionnelles demandées ont été payées dans les délais, le présent rapport de recherche internationale couvre seulement celles des revendications de la demande pour lesquelles les taxes ont été payées, c'est-à-dire les revendications:
- ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale est limité à l'invention mentionnée en premier dans les revendications; elle est couverte par les revendications numéros:

Remarque quant à la réserve

- ☐ Les taxes additionnelles de recherche étaient accompagnées d'une réserve du déposant.
- ☐ Aucune réserve n'a été faite lors du paiement des taxes additionnelles de recherche.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR80/00133

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>3</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. <sup>3</sup> : C 12 N 15/00; C 07 H 21/04; C 12 P 21/02; C 07 C 103/52; A 01 K 39/29; G 01 N 33/54		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>4</sup>		
Classification System	Classification Symbols	
Int.Cl. <sup>3</sup>	C 12 N 15/00; C 12 P 21/00; C 12 P 21/02; C 07 H 21/04	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched <sup>5</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT</b> <sup>14</sup>		
Category <sup>6</sup>	Citation of Document, <sup>15</sup> with indication, where appropriate, of the relevant passages <sup>17</sup>	Relevant to Claim No. <sup>18</sup>
X	Nature, vol. 279, published on 3 May 1979, (USA) C.J. Burrell et al. "Expression in Escherichia coli of hepatitis B virus DNA sequences cloned in plasmid pBR 322.", pages 43 to 47; see figures 1,2,3; page 45, column 1, paragraph 1 to page 47. (Cited in the application)	1 to 4, 8 to 10, 14,15
X	Proc.Natl. Acad. Sci, vol. 76, no.5, published in May 1979, (USA) P.Charnay et al. "Cloning in Escherichia coli and physical structure of hepatitis B virion DNA", pages 2222-2226, see figures 2,3,5; page 2222: "Materials and Methods" (Cited in the application)	1 to 4 8 to 10, 14
X	Nature, vol. 279, published on 24 May 1979, J.J. Sninsky et al. "Cloning and endonuclease mapping of the hepatitis B viral genome", pages 346-348; see the whole document, (Cited in the application)	1 to 4, 8 to 10,14
X	Chemical Abstrats, vol. 86, no.25, published on 20 June 1977, (Columbus, Ohio, USA), D.P. Peterson et al. "Partial amino acid sequence of the major component polypeptides of hepatitis B surface antigen", see page 413 column 2, abstract 187.460r Proc.Natl. Acad.Sci. USA 1977, 74(4). pages 1530-34	15
C	C.R. Acad. Sc. vol. 287, published on 18 December 1978, (Paris, FR)	1
<p><sup>16</sup> Special categories of cited documents:</p> <p>"A" document defining the general state of the art</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document cited for special reason other than those referred to in the other categories</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but on or after the priority date claimed</p> <p>"T" later document published on or after the international filing date or priority date and not in conflict with the application, but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search <sup>19</sup>		Date of Mailing of this International Search Report <sup>20</sup>
26 November 1980 (26.11.80)		19 December 1980 (19.12.80)
International Searching Authority <sup>1</sup> European Patent Office European Patent Office		Signature of Authorized Officer <sup>20</sup>

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, <sup>1a</sup> with indication, where appropriate, of the relevant passages <sup>1c</sup>	Relevant to Claim No <sup>1b</sup>
	Fritsch et al. "Clonage du genome du virus de l'hépatite B dans Escherichia coli", pages 1453-1456, see the whole document (Cited in the application)	-2-
P,X	Nature, vol. 286, published on 28 August 1980, P.Charnay et al. "Biosynthesis of hepatitis B virus surface antigen in Escherichia coli", pages 893-895, see the whole document	14 to 20, 23 to 29
P,X	Nucleic Acids Research, Vol. 7, no. 2, published on 25 September 1979, P.Charnay et al. "Localization on the viral genome and nucleotide sequence of the gene coding for the two major polypeptides of the Hepatitis B surface antigen (HBSAg) pages 335 to 346, see the whole document (Cited in the application)	1 to 13, 15 to 18
P,X	Nature, vol. 281, published on 25 October 1979, F.Galibert et al. "Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E.coli", pages 646-650, see the whole document (Cited in the application)	1 to 13
P,X	Nature, vol. 282, published on 6 December 1979, M.Pasek et al. "Hepatitis B virus genes and their expression in E.coli", pages 575-579, see the whole document, (Cited in the application)	1 to 4, 8, 9, 14, 15
P,X	Nature, vol. 280, published on 30 August 1979, P.Valenzuela et al. "Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen", pages 815-819 (Cited in the application)	1 to 4, 8, 9 14, 15
P,X	E.P. A, 13828, published on 6 August 1980, see page 11, line 1 to page 21, line 3; page 27, line 4 to page 36, claims 1 to 4, 6 to 32; figures 1 to 12, Biogen N.V.	1 to 4, 8 to 10, 14, 15, 19, 20, 23, 25, 26, 27, 29
P,E	DE, A, 2950995, published on 24 July 1980, see page 4, lines 5 to 15; page 6, line 16 to page 17; claims 1 to 12; Institut Pasteur	

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**